FILE COPY FOREST INSECT LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA.

612.
~~032~~2

UNITED STATES DEPARTMENT OF AGRICULTURE

BUREAU OF ENTOMOLOGY AND PLANT QUARANTINE

Project | Date | Author

TITLE

THE CONSTRUCTION OF A CURVE FOR PREDICTING TOTAL ANNUAL LOSS FROM PARTIAL LOSS DATA IN WESTERN PINE BEETLE SURVEYS

By

R. C. Grant, Field Aide,
Bureau of Entomology and Plant Quarantine

April 21, 1941

Forest Insect Laboratory
445 U. S. Court House
Portland, Oregon

JMM
MLL
~~KAG~~
JEP
GMB
HLM
ASW
JWB
CBE
RCH
JSY
STC
PCJ
ESB

SUBJECT—

INDEX NO.—

U. S. GOVERNMENT PRINTING OFFICE 8—9743

THE CONSTRUCTION OF A CURVE FOR PREDICTING TOTAL ANNUAL LOSS FROM PARTIAL LOSS DATA IN WESTERN PINE BEETLE SURVEYS

By

R. C. Grant, Field Aide,
Bureau of Entomology and Plant Quarantine

April 21, 1941

Forest Insect Laboratory
445 U. S. Court House
Portland, Oregon

CONTENTS

	Page
Introduction	1
Subregional Variation	1
New Regional Estimating Curve	6
Some Conditions Affecting the Accuracy of the Estimate	6
Conclusion	8
References	9

LIST OF FIGURES

Figure 1. Percent of Total Annual Loss Spotted on any Date by Subregions

Figure 2. Percent of Total Annual Loss Spotted on any Date for Oregon and Washington.

Figure 3. Comparison of Curves with Those Prepared by Keen and Johnson.

Figure 4. Annual Variation in Percent of Total Annual Loss Spotted on any Date.

Introduction

Men working on surveys of western pine beetle damage in the ponderosa pine region of Oregon, Washington, and California are continually confronted with the problem of predicting the total loss which may develop during the year from partial loss data taken at different times during the season. The first attempt to define this relationship was made in 1931 by F. P. Keen (1). A later analysis was made by P. C. Johnson (2) of conditions found in California. The estimating curves of Keen and Johnson are compared in Figure 3.

Because of the wide variation in accuracy of loss prediction encountered in different years, forest entomologists have expressed some concern about the accuracy of present estimating factors. Since these loss predicting curves were first established, a large amount of additional data have accumulated from the regional ponderosa pine beetle surveys. These data made possible the present study to determine if the present estimating curves can be improved upon and to learn more about some conditions affecting the accuracy of pine beetle loss predictions.

Subregional Variation

The first phase of this study was to determine the accuracy of estimates, for four different subregions, based on local and general estimating factors. Data were taken from 320- and 640-acre pine beetle check plots distributed throughout the ponderosa pine forests of Region 6. These plots have been cruised annually from 1931-39 thus obtaining partial and total annual loss for each year. Surveys were not always initiated on the same dates each year thus making it possible to determine the percent of total annual loss spotted on any date throughout the summer, fall, and winter seasons. A preponderance of mid-season data is available and there have not been enough early and late season cruises to furnish sufficient data for defining the estimating curves at the extremes. But this is not a serious obstacle since recent surveys have been made during mid-season and the estimating percents read from the center of the curve.

The percent of total annual loss spotted on any date was found by dividing the number of trees spotted during the first cruise by the total loss for the year as determined by this and subsequent cruises. The 724 percent determinations were plotted over spotting dates for the following four subregions: Deschutes Drainage, Klamath Basin, Blue Mountains, and Washington. These rough curves served as a means for tabulating the data for comparison of subregional variation and also brought out the great variation in percent of total loss spotted on any one date. The percents were next averaged by ten day classes and put in table form for each subregion and for the entire region. Each table included the date of spotting, mean percent spotted, and the corresponding number of observations. The computed means were then used to draw 5 curves for the region and subregion. These curves showed some differences in slope and shape for each region. Because the curves were plotted graphically and subject to more error than if statistically derived it was decided to determine the regression coefficients for a straight line relationship.

Five correlation tables were set up and the regression of y on x determined, where y = percent of total spotted and x = spotting date. The dates were coded to read as one for June first and numbering consecutively up to 214 for December 31.

The original data all defined curvilinearity and the fitting of a straight line to these data was questionable, so the straight line regressions were corrected to conform more closely to the data. The method used is described by Bruce and Reineke (3). In using this method some allowances were made for frequency distribution. The first step consisted of compiling a table of dates spotted with the corresponding measured mean percents, estimated means from the regression line, and the frequency. The measured values were then plotted over the estimated and a graphic curve drawn. A second table was next made using arbitrarily chosen dates and estimated means with the corresponding measured values read from the graph. The measured values were then plotted over spotting dates and a new regression curve fitted to the points. The above procedure was carried out for the region and four subregions.

The corrected regression lines all showed curvilinearity and there was considerable difference in the shapes and slopes of the curves for the subregions. For a comparison of these curves refer to Figure 1.

The ultimate aims in the development of these curves were to obtain as low individual variations as possible between predicted and actual number of trees killed and a low variation in the gross estimate for large areas over a period of years on the sample plots. In order to determine the accuracy of the various estimating curves the actual and predicted 1938 and 1939 insect losses in number of trees were compared on several sample plots in Oregon and Washington. For each subregion three estimates were made, as follows: one from the estimating curve now in use, another from the regional curve, and estimates from the corresponding subregional curve.

The standard errors of the estimate and the percentages of aggregate differences of estimated from actual number of trees killed were used to check the relative accuracy of the curves. The percent of aggregate difference is a measure of the percent difference between total number of trees estimated and the total number actually killed. The standard error of the estimate, which is the root mean square of the deviations from the line of estimate, measures the number of trees the individual loss estimates will vary from the actual loss 68 percent of the time.

The two measures of the estimate act as checks on each other. An estimate may have a low aggregate difference with a high standard error; and in such a case it is likely that the errors of estimate were high but compensated each other thus giving a low aggregate difference. In those cases where the standard errors of the estimate and the aggregate differences are both low the estimates can be considered reliable. Both measures of accuracy showed a slight advantage to lie with the estimates made from curves based on subregional or local data for the area estimated. But it was decided to use the regional curve since the simplicity and ease in using one curve offsets the very slight increase in accuracy obtained by using four local curves. The following tabulation gives a comparison of the three estimates by estimating from each curve the total annual loss in eastern Oregon and Washington and determining the percent of difference from the total number of trees actually killed.

	Percent of Aggregate Difference from Actual Number of Trees Killed		
	1938	1939	Mean
Old Regional Estimate	+15.7	+11.9	13.80
New Regional Estimate	+ 6.5	- 5.8	6.15
Subregional Estimate	+ 4.1	- 8.0	6.05

Table I gives a detailed account of the accuracy shown by the three estimating curves for the four regions in 1938 and 1939. Washington had the lowest standard error in 1938 when the estimated expectancy 68% of the time was 7.4 trees from the actual number killed. The Deschutes was high in 1939 with a standard error of 38.8 trees.

If in checking these curves, we used samples representative of future conditions, we can assume that our future predictions of total annual loss made from the new regional curve will approximate the accuracy shown for this curve in Table I.

Table I.—Comparison of Three Estimates of Total Trees Killed on Four Subregions in Oregon and Washington.

Subregion	Percent Aggregate Difference of Estimated from Actual Number Trees Killed: Old Estimate Percent	Regional Estimate Percent	Local Estimate Percent	Standard Error of Estimate in Number of Trees: Old Estimate Trees	Regional Estimate Trees	Local Estimate Trees
Year 1938						
Blue Mountains	+16.9	+11.7	+ 9.6	15.7	13.9	12.8
Klamath Basin	- 8.5	- 3.5	- 0.8	22.4	23.2	22.4
Deschutes	+54.2	+11.5	- 1.6	37.5	20.3	18.0
Washington	+ 2.5	+ 3.4	+ 8.9	9.4	7.4	7.5
Year 1939						
Blue Mountains	+ 3.3	+ 2.5	+ 4.2	13.9	11.2	11.2
Klamath Basin	+21.2	+ 7.4	+ 9.9	20.0	13.3	14.5
Deschutes	+34.6	-14.9	-26.2	38.8	26.1	28.4
Washington	-21.8	-23.6	-21.0	13.7	13.9	13.0

In the foregoing phase of this analysis we have eliminated latitudinal or subregional variations as strong factors affecting estimates of total annual loss. We may, without further statistical analysis, eliminate elevation as an important variable since the two, elevation and latitude, are closely correlated in their effects on plant and animal life. Although there are differences in the seasonal length and inception of bark-beetle attacks at varying elevations the practicality of measuring these changes is doubtful since adjustments for elevation would have to be made commensurately with latitudinal changes. The increase in accuracy resulting from the inclusion of elevation as a variable would probably approach that obtained by the addition of latitude as a variable, and such a refinement is not justifiable. The spread of the percent of total loss spotted on any one date ranges from 10 to 70 percent and an estimate at its best, made from these data will have a high standard error. In analyzing highly variable data of this nature it is often the tendency to add refinements until the worker ends in confusion with so many variables that it is safer to hazard a guess.

New Regional Estimating Curve

After weighing all possible factors against the highest accuracy we could expect from the data at hand, a new regional curve was constructed. This curve was based on survey data covering a 17-year period over which 1,153 percent determinations were computed; and should serve as a good basis for a curve applicable to all general conditions (Fig. 2). It may be necessary to alter the curve for closer conformity to local conditions or a seasonal trend but it serves as a good basis from which to start. By using the method previously described for curving a straight line to fit curvilinear data it is also possible to adjust the general curve to fit data more representative of the locality or the trend. The shapes and slopes of the curves in Figure 3 give an indication of the extent of the adjustments which might be made to the normal curve. In these instances adjustments would be made for periodic and subregional variations.

Some Conditions Affecting the Accuracy of the Estimate

As has been brought out already, there are many factors which affect the estimate of total annual loss. Some of the factors mentioned by Keen (1) were the personal equation of the spotters, the method of cruising used, errors due to time of year and those due to variations between seasons, and errors due to sudden changes in the infestation trend. Johnson (2) discusses the effect on accuracy of estimates of

such factors as trend and composition of infestation, composition of stand, climatic conditions, elevation and variation in spotting. A statistical analysis of such a variety of variables would demand an adequate distribution of plots in each of the cases and is well outside the scope of present available data. In fact, the extreme variability of the basic data would preclude any attempt to consider these other variables and moreover some of these factors do not lend themselves to mathematical determination.

While some factors are outside the realm of statistical analysis, under the present status of available information, a brief consideration of them may be of help to the estimator. Climate produces varied effects on the inception and progression of attacks, the extent of which cannot be measured for use in a curve but should act as a guide to an experienced estimator. As a general rule it may be stated that any climatic factor which increases the resistance of trees to beetle attack or retards beetle activity will result in an overestimate of total annual loss. Losses are overestimated to some extent when the active season is preceded by very low winter temperatures. Keen (1) shows a relationship between number of clear days and the estimate. A large number of clear days during the year may cause an underestimate of beetle damage. Latitude and elevation changes produce effects only in so far as favorable growing temperatures are retarded or advanced. Adjustments for latitude, elevation, cold basins and other local differences should be made by the observer.

A study of the annual variations in percent of total loss spotted on any date was made for the Klamath Subregion during the 17-year period 1923-39. The regression coefficients for each year were determined and the straight lines of closest fit plotted on one graph to show the variation between years (see Figure 4).

As a general rule total annual loss is apt to be underestimated during periods of increasing beetle activity but this is dependent on the season during which the estimate is made. Epidemics usually progress irregularly and total annual loss is difficult to estimate during early season spotting. It is quite possible that estimates made from a normal curve may be low during the early part of the season and high at the end or vice versa. In such cases the errors of estimate may be compensating for the year. A knowledge of the effects caused by sudden changes in the infestation and the trend of an epidemic should be more helpful to the estimator during epidemics than estimating coefficients read from a normal curve of total loss spotted on any date.

Since estimating loss for the year is based on a percent of the trees already spotted by a crew, it is reasonable that some allowances should be made for attacked trees which have not been detected by the spotters. No allowances need be made for a crew of average spotters because the curves in this report are based on work done under average conditions. But some allowances should be made for crews either above or below the average. Information gathered from studies of control work shows that spotting crews under favorable conditions normally detect about 90 percent of current attacks.

Spotting has generally been done during August and September on the pine beetle surveys, and three conditions, as far as estimating total annual loss is concerned, favor the present practice of making mid-season surveys, namely: the strength of the estimating curve during this period, the small annual variation, and the small sub-regional variation during mid-season. (Refer to Figures 1 and 4.) The high standard error of the estimate of early season spotting is shown in Table I by the Deschutes Subregion, which was spotted in June and July for the two years represented in the table.

Conclusion

A new curve for predicting total annual loss from partial loss data taken during the active period from June 1 to December 20 has been constructed from a statistical analysis of 1,153 observations on sample plots cruised in the ponderosa pine region of eastern Oregon and Washington during the 17-year period 1923 to 1939. This new curve is shown in Figure 2. It is expected that this curve will give a better estimate of total annual loss than the one previously used.

Differences in four principal subregions were tested and while slight improvement in estimates was indicated, this was not sufficiently great to justify the inconvenience of using four local curves in place of one regional curve.

Klamath Subregion data showed considerable annual variation in percent spotted on any one date. Annual variation is largely a function of climatic changes and beetle trends, and no correction for these is proposed other than the entomologist's ability to predict the effect of weather and the trend of beetle activity from his general observations and past experience.

References

1. Keen, F. P. Estimating Pine Beetle Losses During Period of Attack. Portland Station Report. Feb. 6, 1931.

2. Johnson, P. C. Correction Factors for Use in Estimating Total Seasonal Loss. Berkeley Station Report. Feb. 5, 1935.

3. Bruce, Donald, and Reineke, L. H. Correlation Alinement Charts in Forest Research. U.S.D.A. Tech. Bull. No. 210. February 1931.

FIGURE 1

PERCENT OF TOTAL ANNUAL LOSS SPOTTED ON ANY DATE

OREGON AND WASHINGTON

1931 - 1939

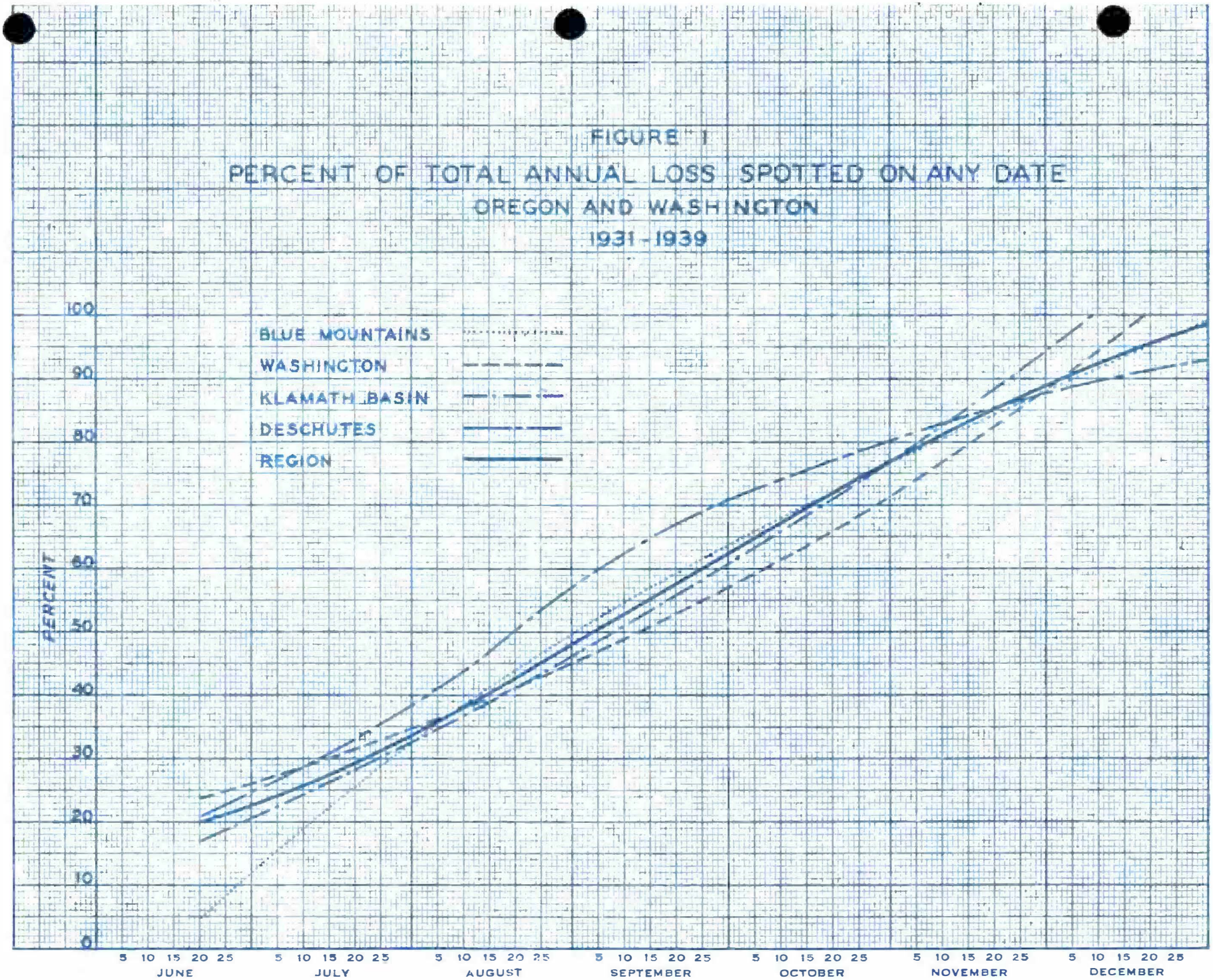


FIGURE 2

PERCENT OF TOTAL ANNUAL LOSS SPOTTED ON ANY DATE

OREGON AND WASHINGTON PERIOD OF 1923-1939

1153 OBSERVATIONS

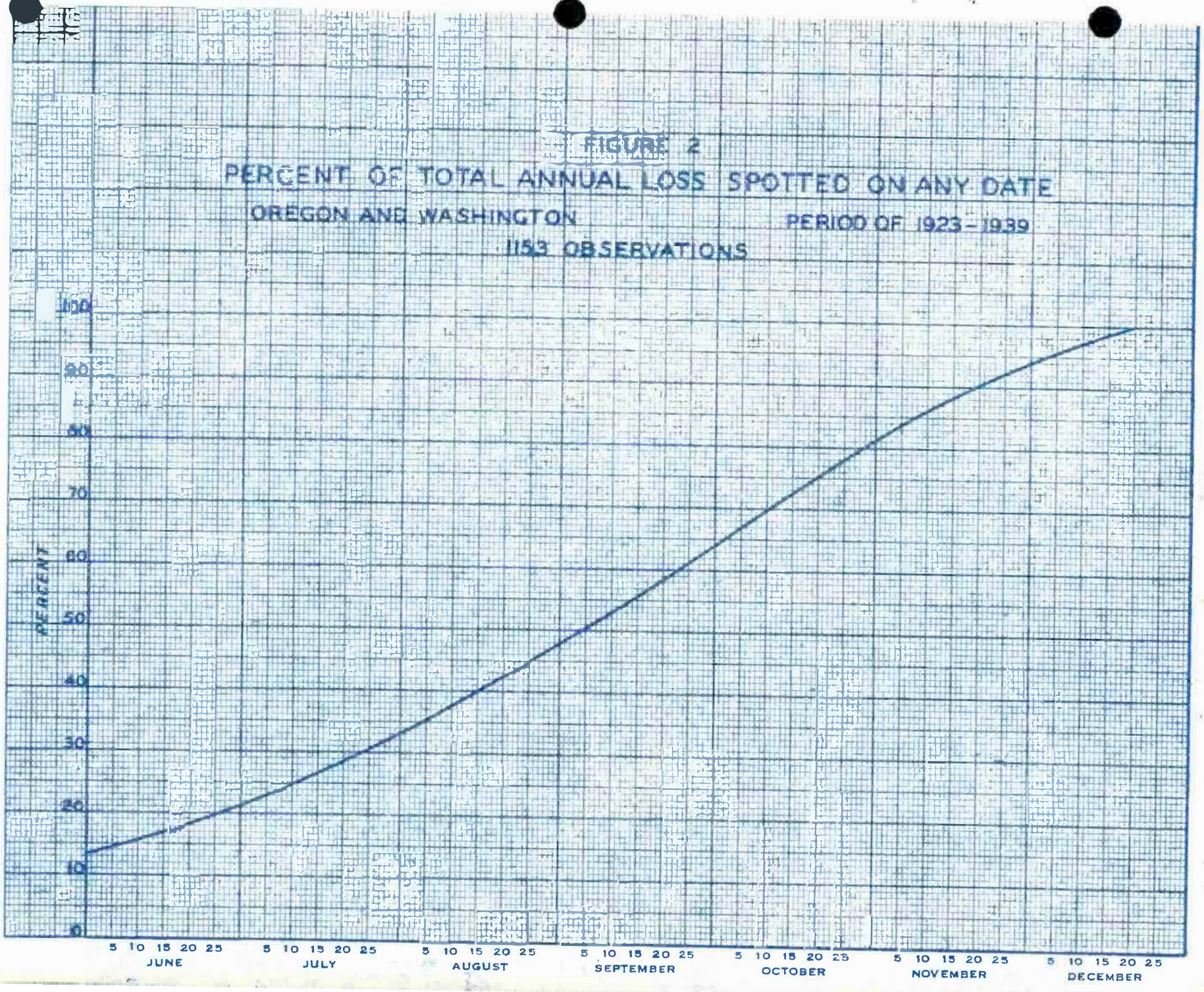
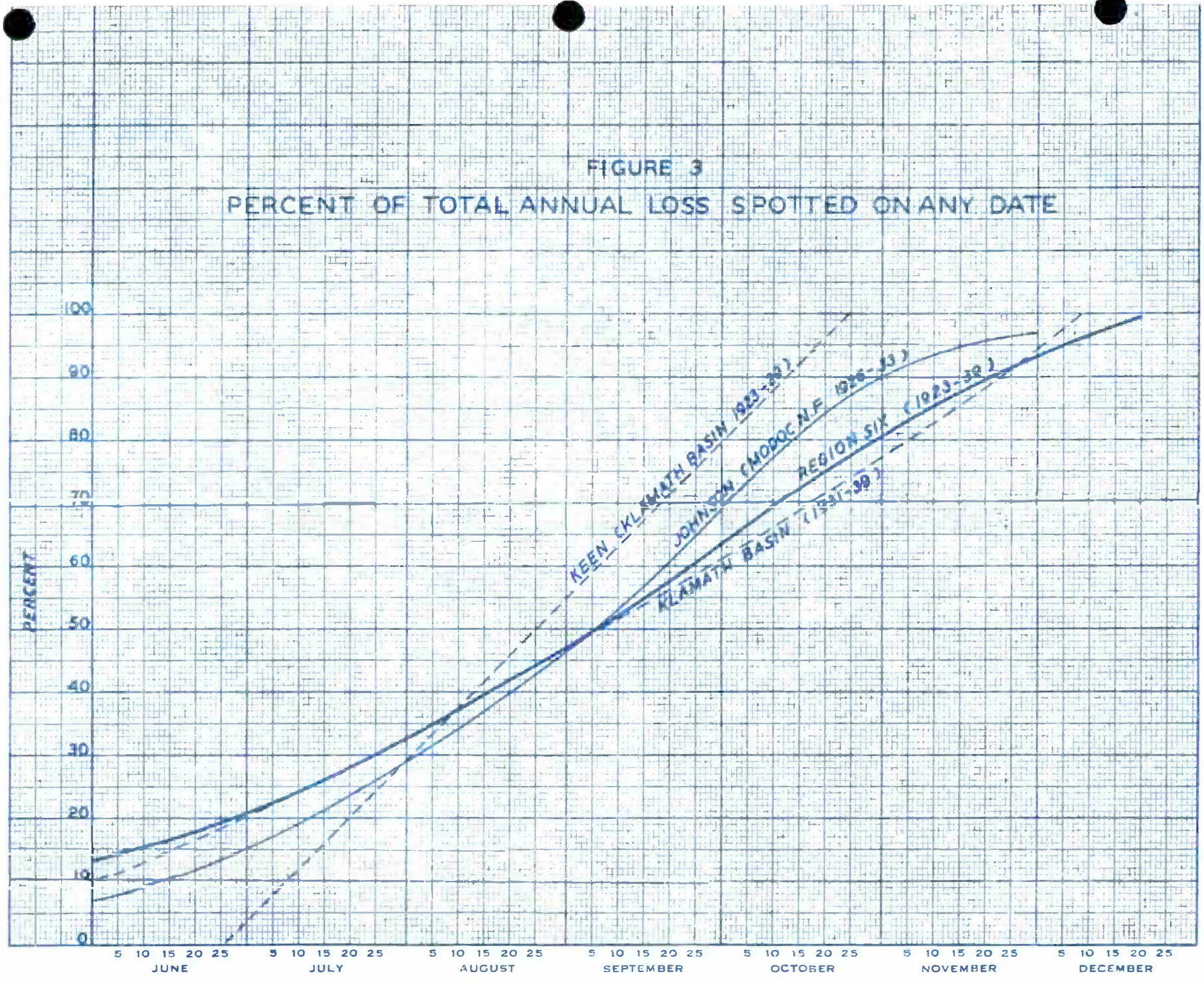


FIGURE 3

PERCENT OF TOTAL ANNUAL LOSS SPOTTED ON ANY DATE



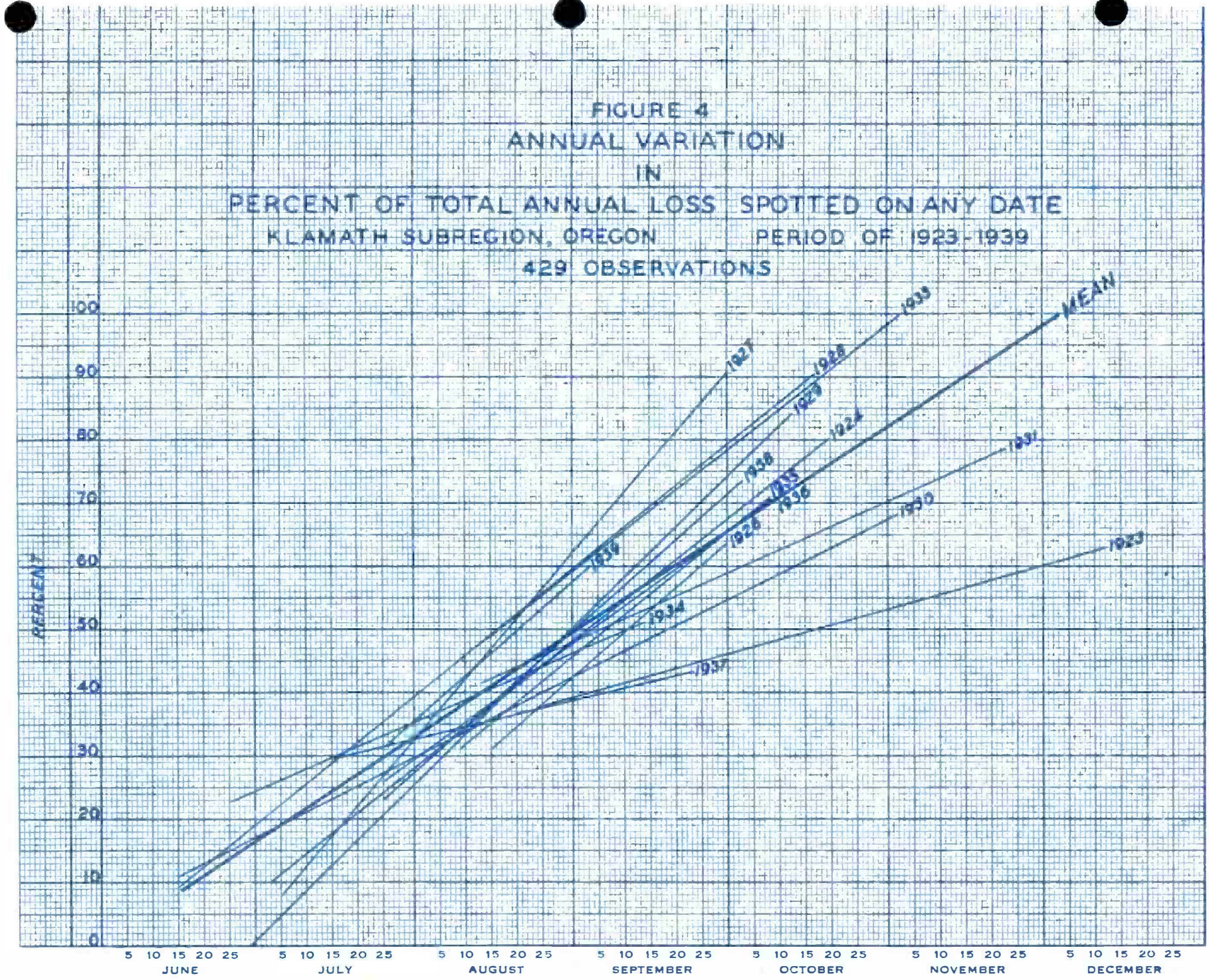


FIGURE 4
ANNUAL VARIATION
IN
PERCENT OF TOTAL ANNUAL LOSS SPOTTED ON ANY DATE
KLAMATH SUBREGION, OREGON PERIOD OF 1923-1939
429 OBSERVATIONS
PERCENT
100
90
80
70
60
50
40
30
20
10
0
MEAN
1933
1927
1928
1929
1924
1931
1938
1935
1936
1930
1926
1923
1939
1934
1937
5 10 15 20 25
JUNE
JULY
AUGUST
SEPTEMBER
OCTOBER
NOVEMBER
DECEMBER